

DESENVOLVIMENTO DE UMA PLATAFORMA COMPUTACIONAL PARA OBTENÇÃO DA FORMA 3D DE OBJECTOS USANDO TÉCNICAS DE VISÃO ACTIVA

Teresa Azevedo¹, João Manuel R. S. Tavares^{1,2}, Mário A. Vaz^{1,2}

¹Instituto de Engenharia Mecânica e Gestão Industrial, Lab. Óptica e Mecânica Experimental
e-mail: {teresa.azevedo, tavares, gmavaz}@fe.up.pt

²Universidade do Porto, Faculdade de Engenharia, Dep. de Engenharia Mecânica e Gestão Industrial
e-mail: {tavares, gmavaz}@fe.up.pt

RESUMO

Neste artigo pretende-se descrever uma plataforma computacional que está a ser desenvolvida para obter a forma 3D de objectos usando técnicas de Visão Activa. Assim, partindo-se de uma sequência de imagens não calibradas do objecto a reconstruir, usando a referida plataforma, pretende-se obter a geometria 3D do objecto em causa.

1 – INTRODUÇÃO

Visão Computacional é a área da ciência que se dedica a desenvolver teorias e métodos de extracção automática de informações úteis em imagens, de forma o mais semelhante possível à realizada pelo ser humano no seu complexo sistema visual.

Nesta área são várias as técnicas sem contacto que, directa ou indirectamente, se podem utilizar para recuperar a geometria 3D de um objecto, a distância desse objecto a um determinado referencial ou a orientação local da sua superfície. Tendo em conta os princípios básicos utilizados, essas técnicas são normalmente divididas em passivas e activas (Fig. 1).

O principal objectivo do trabalho inerente a este artigo é a obtenção de modelos 3D de objectos utilizando a metodologia usualmente designada por *Reconstrução de Estruturas a partir do Movimento* (na literatura anglo-saxónica, *SFM – Structure from Motion*), que se insere no grupo das técnicas de Visão Activa (Pollefeys, 1998).

Apresentada pela primeira vez em 1979 (Ulman, 1979), a técnica *SFM* recebeu, ao longo do tempo, contribuições e abordagens diversificadas. Dentro das

possibilidades de variação, salienta-se o número de vistas distintas consideradas do objecto, a utilização ou não de pontos fortes e o conhecimento prévio ou não das características do sistema de aquisição e o posicionamento relativo da(s) câmara(s) utilizadas.

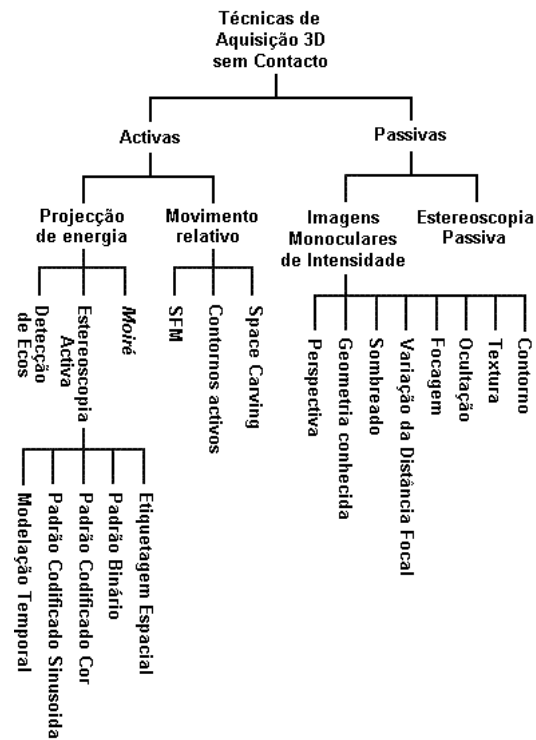


Fig. 1 - Divisão usual das técnicas sem contacto para obtenção de informação tridimensional de um objecto.

No caso presente, não se pretende impor qualquer tipo de restrição ao movimento relativo envolvido. Assim, partindo de uma sequência de imagens não calibradas sobre um objecto, pretende-se extrair o movimento em causa, a calibração da(s) câmara(s) considerada(s), e, finalmente, obter a geometria 3D do objecto em causa.

De forma a conseguir o objectivo proposto, está em desenvolvimento uma plataforma computacional na qual estão a ser integradas diversas funções disponibilizadas em várias bibliotecas de domínio público.

As funções já integradas abrangem várias técnicas de Visão Activa: extracção e emparelhamento de pontos fortes (pontos característicos) entre imagens, determinação da geometria epipolar, rectificação e emparelhamento denso (Azevedo, 2005). Estas técnicas são habitualmente usadas na obtenção da forma 3D de objectos a partir de sequências de imagens (Fig. 2).

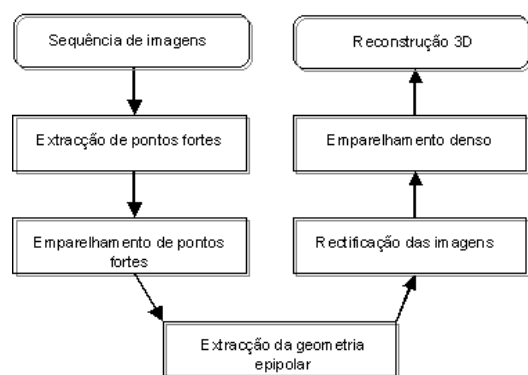


Fig. 2 - Estrutura da metodologia adoptada neste trabalho para a reconstrução 3D de objectos a partir de sequências de imagens.

Neste artigo, após uma breve descrição da plataforma computacional em desenvolvimento, são apresentados alguns resultados experimentais obtidos e, no final, são enumeradas algumas conclusões e indicadas perspectivas de trabalho futuro.

2- PLATAFORMA DE VISÃO ACTIVA

A plataforma computacional está a ser desenvolvida usando a linguagem C^{++} , com uma interface gráfica adequada para o

utilizador, onde se integraram algumas das funções disponibilizadas em cinco *softwares* e uma biblioteca computacional, todos de domínio público (sobre os fundamentos das funções integradas ver Azevedo, 2005), para a reconstrução de objectos usando técnicas de Visão Activa. Tais entidades serão neste artigo referidas genericamente como *Programas 1 a 6*:

- *Programa 1 – Peter’s Matlab Functions for Computer Vision and Image Analysis* (Kovesi, 2004): pacote de funções em *Matlab* de processamento e análise de imagem e Visão 3D;
- *Programa 2 – Torr’s Matlab Toolkit* (Philip Torr, 2002): com uma interface gráfica em *Matlab*, aplica a técnica *SFM* entre duas imagens;
- *Programa 3 – OpenCV* (2004): biblioteca de funções em C^{++} que implementam alguns dos algoritmos mais usuais no domínio da Visão Computacional;
- *Programa 4 – Kanade-Lucas-Tomasi (KLT) Feature Tracker* (Birchfield, 2004): conjunto de funções em C , que implementam o algoritmo de *KLT*, para extrair e emparelhar pontos numa sequência de imagens (B. D. Lucas, 1981, C. Tomasi, 1981);
- *Programa 5 – Projective Rectification without Epipolar Geometry* (F. Isgrò, 1999): programa em C , que realiza a rectificação de duas imagens estéreo sem predeterminar a geometria epipolar;
- *Programa 6 – Depth Discontinuities by Pixel-to-Pixel Stereo* (Birchfield, 1999): programa em C , que retorna os mapas de disparidade e descontinuidades entre duas imagens rectificadas.

Os *programas* desenvolvidos originalmente na linguagem *Matlab* foram, neste trabalho, portados para C , de forma a serem convenientemente integrados na plataforma computacional.

Na referida plataforma, para cada técnica de Visão Activa disponibilizada, o utilizador pode facilmente escolher o algoritmo que pretende usar (Fig. 3), assim

como definir comodamente os seus parâmetros (Fig. 4).

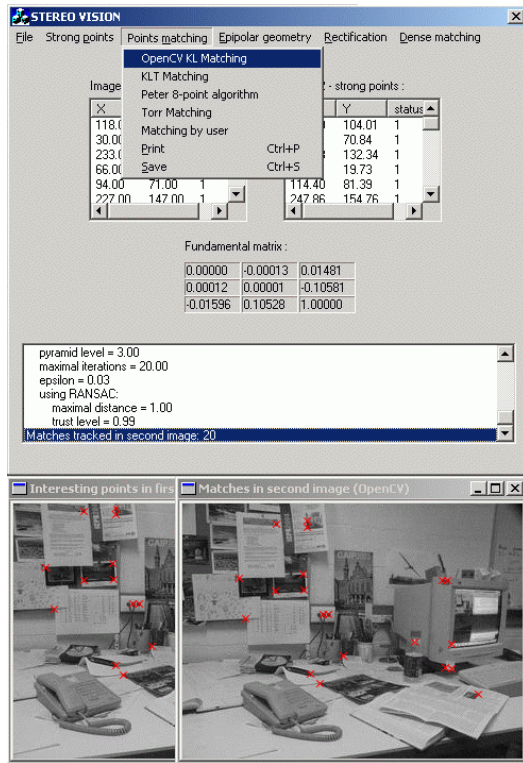


Fig. 3 - Em cada item do menu da plataforma, são apresentados os vários algoritmos disponíveis para cada técnica; em baixo, duas imagens resultantes da aplicação do algoritmo de emparelhamento de pontos fortes do *programa OpenCV*.

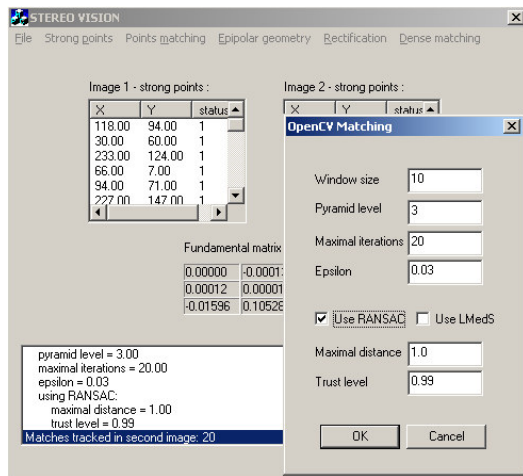


Fig. 4 - Exemplo de selecção dos parâmetros do algoritmo de emparelhamento de pontos fortes do *programa OpenCV*, disponível na plataforma computacional.

3 – RESULTADOS EXPERIMENTAIS

Nos testes experimentais, utilizaram-se pares de imagens estéreo, de dimensões 540×612 píxeis, obtidas de vários

objectos reais, capturadas com uma câmara digital convencional.

Alguns dos resultados experimentais obtidos são apresentados nesta sessão segundo a sequência da Fig. 2.

3.1 – Extracção de pontos fortes

O primeiro passo da maioria das técnicas usuais de Visão 3D é a detecção de pontos fortes numa imagem. Esses pontos são aqueles que têm uma componente 2D característica sobre o objecto; ou seja, que reflectem discrepâncias relevantes entre os seus valores de intensidade e os dos seus vizinhos. Usualmente, estes pontos são os vértices do objecto, e a sua extracção permite posteriormente correlacioná-los noutras imagens da sequência. Este emparelhamento torna possível a posterior reconstrução 3D do objecto em causa.

A título de exemplo, na Fig. 5 são apresentados os resultados obtidos pelos *programas 1 a 4*, na extracção de pontos fortes numa das imagens de teste consideradas. Como se pode observar, o *programa 2* é o que apresenta resultados mais fracos, principalmente pelo aparecimento de zonas com uma densidade muito elevada de pontos característicos. Tal deve-se ao facto da técnica em causa não garantir uma distância mínima entre estes pontos, tal como acontece com as técnicas consideradas nos *programas 3 e 4*.

3.2 – Emparelhamento de pontos fortes

Na reconstrução 3D é necessário associar os pontos 2D nas várias imagens da sequência, que resultem da projecção do mesmo ponto da cena (*matching*). Geralmente, um reduzido número de pontos de correspondência é suficiente para se poder determinar uma relação geométrica entre duas imagens; ou seja, para se obter a matriz fundamental.

A detecção automática das correspondências existentes entre os pontos fortes de duas imagens parcialmente sobrepostas, pode ser realizada através do cálculo de uma das várias medidas de *matching* existentes.

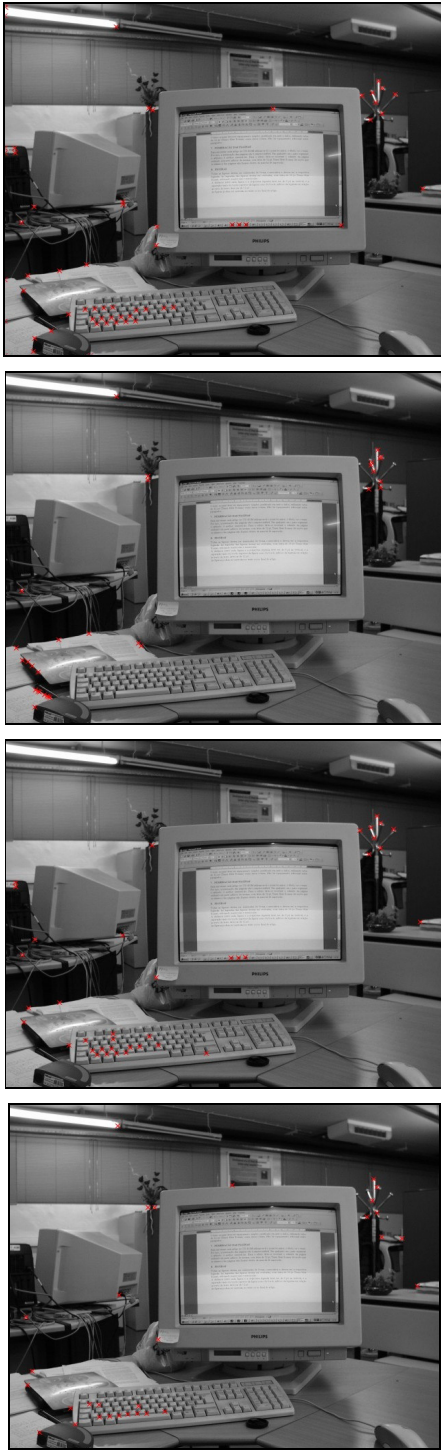


Fig. 5 - Resultado da extracção de pontos fortes determinados numa das imagens do par estéreo considerado: as cruces a vermelho indicam os pontos fortes encontrados (de cima para baixo, são os resultados dos programas 1 a 4).

Geralmente, essas medidas, para cada ponto forte numa imagem I , calculam o grau de semelhança em toda uma imagem J ou para todos os pontos fortes da imagem J , sendo para isso

analisada e comparada a vizinhança de cada um desses pontos.

Após a determinação dos pontos fortes na primeira imagem, as funções de emparelhamento tornam-se disponíveis.

Exemplos de resultados experimentais obtidos pelas técnicas integradas na plataforma são apresentados na Fig. 7. Nestes, observa-se que o emparelhamento de pontos fortes é uma fase crítica em Visão 3D. Em todos os algoritmos considerados, verificam-se erros na correlação, sendo o *programa 2* o que apresenta os resultados mais fracos.

3.3 – Determinação da geometria epipolar

Dado um par de imagens estéreo, qualquer ponto P do espaço 3D define um plano π_p que contém esse ponto e os centros ópticos O_l e O_r de duas câmaras. O plano π_p referido é denominado de *plano epipolar*, e as linhas de intersecção deste plano com os planos 2D imagem de cada câmara são designadas por *linhas epipolares conjugadas*, Fig. 6.

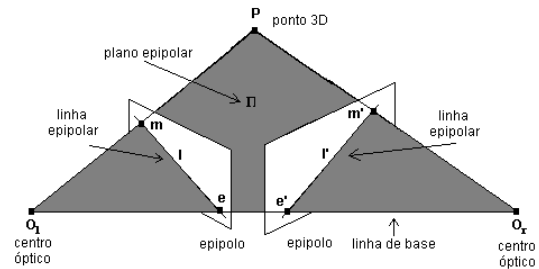


Fig. 6 – Geometria epipolar.

A geometria epipolar corresponde à estrutura geométrica entre as duas vistas e expressa-se matematicamente pela matriz fundamental F , de dimensão (3×3) :

$$m^T F m = 0,$$

onde m e m' são os pontos de emparelhados das duas imagens. Cada ponto m , está sobre a linha epipolar l respectiva, definida por:

$$l = F m'.$$

A determinação da geometria epipolar significa obter informação da pose

relativa entre duas vistas da mesma cena. Tal informação permite também eliminar erros no emparelhamento efectuado previamente, assim como facilitar a posterior obtenção de novos pontos de correlação (fase de emparelhamento denso).

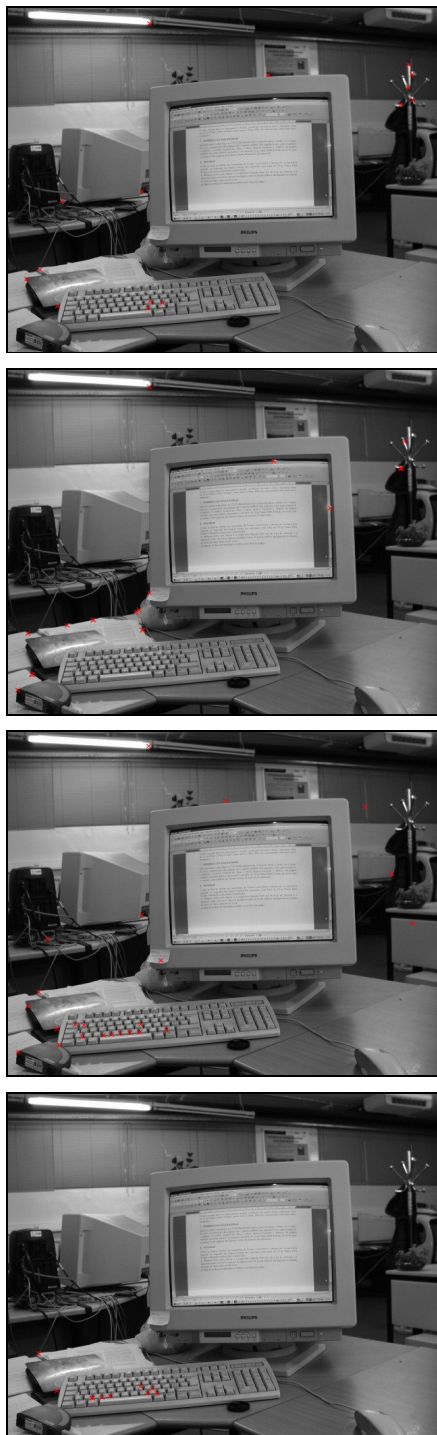


Fig. 7 - Resultado do emparelhamento dos pontos fortes, da Fig. 5, na outra imagem do par estéreo considerado: as cruzes a vermelho indicam os pontos emparelhados (de cima para baixo, são os resultados dos programas 1 a 4).

Os resultados experimentais obtidos, a título de exemplo, são apresentados na Fig. 8.

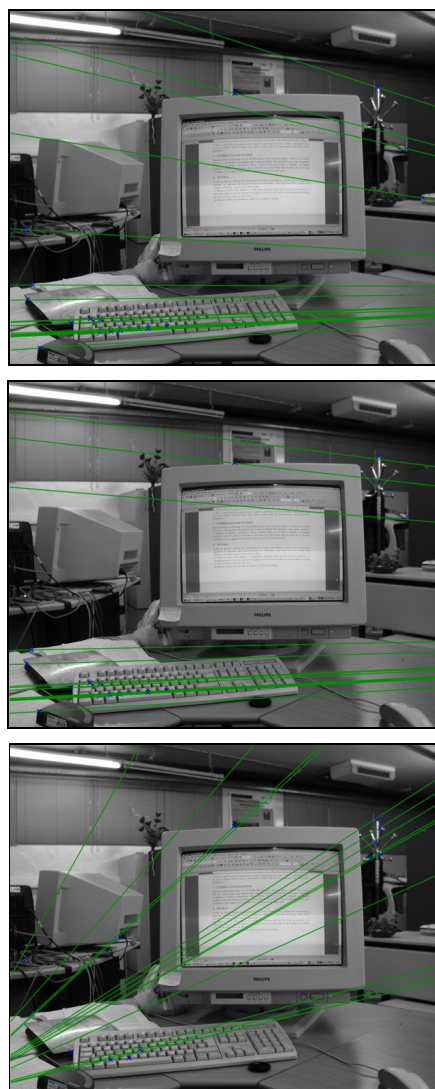


Fig. 8 - Linhas epipolares (a verde) e *inliers* (a azul) obtidos na primeira imagem do par considerado, após a determinação da geometria epipolar; de cima para baixo, resultado do *programa 1* pelo método de Ransac (M. Fischler, 1981), resultado do *programa 1* pelo método de LmedS (R. J. Rousseeuw, 1986), e resultado do *programa 2*.

Como os resultados de fraca qualidade apresentados pelo *programa 2* nas sub-secções anteriores faziam prever, não foi possível obter a geometria epipolar utilizando os algoritmos por este considerados, visto o número de *outliers* – pontos incorrectamente emparelhados substancialmente superior aos *inliers* – pontos correctamente emparelhados.

Dos restantes *programas*, o que aparenta melhores resultados é o *programa 3* (*OpenCV*), embora se note a necessidade

de obter-se melhores pontos de emparelhamento no sentido de conseguir otimizar os resultados obtidos na estimativa da geometria epipolar.

3.4 – Rectificação

Se as imagens forem alteradas de forma a ficarem coplanares a um plano comum, as linhas epipolares tornam-se paralelas ao eixo horizontal da imagem: operação designada por rectificação. Desta forma, torna-se mais fácil a posterior resolução do problema de emparelhamento denso entre imagens.

O *programa 5* é o único a realizar a operação de rectificação, sem para isso ser necessário a determinação da geometria epipolar, ou mais especificamente, da matriz fundamental. Os resultados obtidos pelo programa referido estão apresentados na Fig. 9.

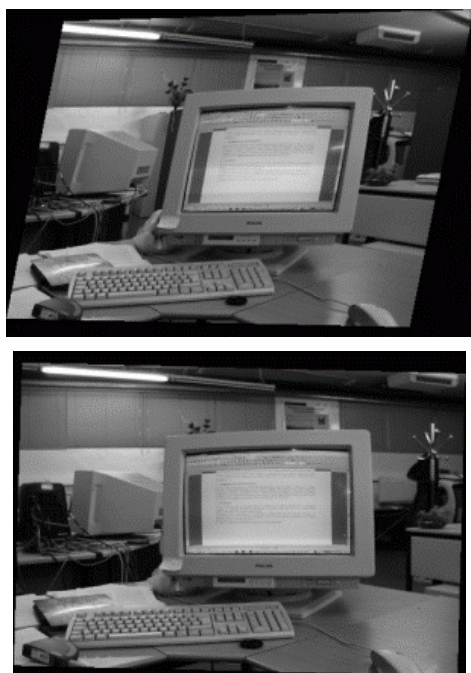


Fig. 9 - Resultado da rectificação pelo *programa 5* das imagens de teste consideradas.

Como se pode observar nas duas imagens resultantes apresentadas, a qualidade dos resultados obtidos, é proporcional à qualidade dos resultados conseguidos na determinação da geometria epipolar.

3.5 – Mapa de disparidade

Num sistema estéreo, a informação 3D de uma cena é representada pela disparidade entre duas imagens. O mapa de disparidade codifica a distância dos objectos em relação à(s) câmara(s); ou seja, pontos muito distantes têm disparidade zero (geralmente 0, equivalente ao preto) e pontos muito próximos terão a máxima disparidade (usualmente 255, correspondente ao branco). Em resumo, um mapa de disparidade dá a percepção das descontinuidades em termos de profundidade de uma cena.

Somente os *programas 3* e *6*, realizam esta operação, com a diferença de que o *programa 6* também retorna um mapa de descontinuidades, definido pelos píxeis que fazem fronteira entre uma mudança de pelo menos dois níveis de disparidade.

Esta etapa foi testada com imagens, previamente rectificadas e incluídas no *programa 6* (Fig. 10). Os resultados obtidos estão apresentados nas Fig. 11 e Fig. 12.

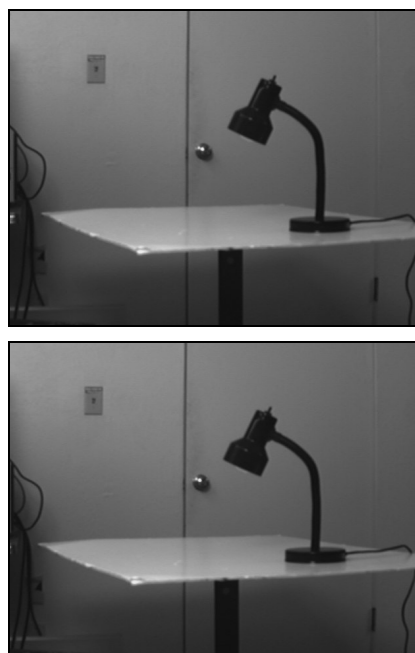


Fig. 10 - Imagens originais utilizadas pelos *programas 3* e *6* para obter o mapa de disparidade (associado ao emparelhamento denso).

Como se pode observar pelos resultados experimentais apresentados,

tanto o *programa 3* como o *programa 6*, obtêm bons resultados na determinação do mapa de disparidade, dadas duas imagens estéreo adequadamente rectificadas.

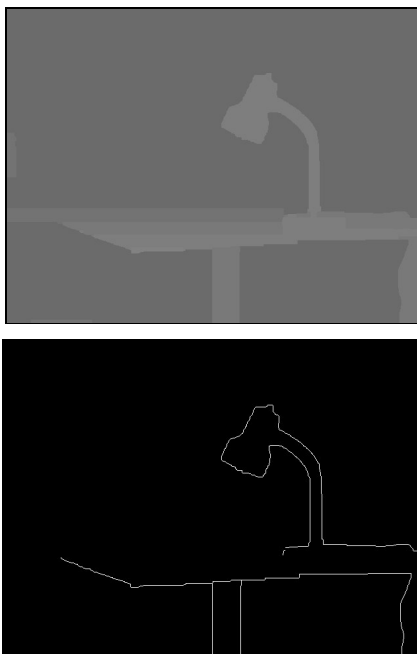


Fig. 11 - Mapas de disparidade (em cima) e descontinuidade (em baixo) obtidos pelo *programa 6*.



Fig. 12 - Mapa de disparidade obtido pelo *programa 3*.

4– CONCLUSÕES

Ao longo deste trabalho, iniciou-se o desenvolvimento de uma plataforma computacional, com uma interface gráfica adequada, para permitir a aplicabilidade e a comparação de várias funções de Visão 3D para a reconstrução 3D de objectos.

Em resumo, pode-se afirmar que as funções, integradas na referida plataforma e experimentalmente analisadas, obtêm bons resultados quando aplicadas em objectos que apresentam características fortes. Dos resultados experimentais

obtidos, é possível concluir que os resultados de menor qualidade estão fortemente relacionados com a determinação dos pontos relevantes nas imagens a considerar e do emparelhamento obtido entre os mesmos. Assim, como as funções a seguir usadas para obter a reconstrução 3D utilizam estes pontos relevantes, os resultados obtidos pelas mesmas estão fortemente dependentes da sua qualidade.

5– TRABALHO FUTURO

As próximas etapas deste trabalho, irão concentrar-se no sentido de melhorar os resultados obtidos quando os objectos a reconstruir apresentam formas contínuas e suaves. Assim, a abordagem em perspectiva será a seguinte:

- consideração de técnicas de “escavação” do espaço tridimensional – *space carving* – para a reconstrução do objecto (ver, por exemplo, [S. N. Seitz, 1997], [K. N. Kutulatos, 1998], [M. Sainz, 2002], [A. Montenegro, 2004]);
- os pontos característicos a usar na definição do espaço 3D “a escavar” serão identificados utilizando um reduzido número de marcadores nos objectos;
- integração de uma técnica de calibração da(s) câmara(s) utilizada(s), assim como algoritmos de obtenção da pose e do movimento da(s) mesma(s); as técnicas base a considerar são as descritas em [Meng, 2000], [Zhang, 1996], [Zhang, 2000];

Posteriormente, pretende-se utilizar a plataforma computacional apresentada na reconstrução e caracterização de formas 3D anatómicas exteriores.

REFERÊNCIAS

- [1] M. Pollefeys, L. V. Gool, M. Vergauwen, R. Koch, *Flexible acquisition of 3D structure from motion*, Proceedings of IMDSP'98 - 10th IEEE Workshop on Image and Multidimensional Digital Signal Processing, Alpbach, Austria, pp. 195-198, 1998.

- [2] S. Ulman, *The Interpretation of Visual Motion*, Mass. MIT Press, Cambridge, 1979.
- [3] T. Azevedo, J. R. S. Tavares, M. P. S. Vaz, *Obtenção da Forma 3D de Objectos usando Metodologias de Reconstrução de Estruturas a partir do Movimento*, Congreso de Métodos Numéricos en Ingeniería, Granada, Espanha, 2005.
- [4] P. Kovesi, *MATLAB Functions for Computer Vision and Image Analysis*, <http://www.csse.uwa.edu.au/~pk/Research/MatlabFns>, 2004.
- [5] Philip Torr, *A Structure and Motion Toolkit in Matlab*, <http://www.cms.brookes.ac.uk/~philip/Tools/Beta/torrsm.zip>, 2002.
- [6] *OpenCV - Open Computer Vision Library*, beta 4, <http://sourceforge.net/projects/opencvlibrary>, 2004.
- [7] S. Birchfield, *KLT: An Implementation of the Kanade-Lucas-Tomasi Feature Tracker*, Stanford Vision Lab, Stanford University, <http://www.ces.clemson.edu/~stb/klt/index.html>, 2004.
- [8] T. K. B. D. Lucas, *An Iterative Image Registration Technique with an Application to Stereo Vision*, International Joint Conference on Artificial Intelligence, pp. 674-679, 1981.
- [9] T. K. C. Tomasi, *Detection and Tracking of Point Features*, Carnegie Mellon University Technical Report CMU-CS-91-132, 1981.
- [10] E. T. F. Isgrò, *Projective rectification without epipolar geometry*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR99, Fort Collins (Colorado), vol. I, ISBN 0-7695-0149-4, pp. 94-99, 1999.
- [11] S. Birchfield, *Depth Discontinuities by Pixel-to-Pixel Stereo*, International Journal of Computer Vision, vol. 35, no. 3, pp. 269-293, <http://vision.stanford.edu/~birch/p2p/>, 1999.
- [12] R. B. M. Fischler, *RANdom SAMpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography*, Commun. Assoc. Comp. Mach., Vol. 24, No. 6, pp. 381-395, 1981.
- [13] A. M. L. R. J. Rousseeuw, *Robust Regression and Outlier Detection*, John Wiley & Sons, 1986.
- [14] C. R. D. S. N. Seitz, *Photorealistic Scene Reconstruction by Voxel Coloring*, Proc. Computer Vision and Pattern Recognition Conference, pp. 1067-1073, 1997.
- [15] S. M. S. K. N. Kutulatos, *A Theory of Shape by Space Carving*, Technical Report TR692, Computer Science Dept., U. Rochester, 1998.
- [16] N. B. M. Sainz, A. Susin, *Carving 3D Models from Uncalibrated Views*, 5th IASTED International Conference Computer Graphics and Imaging (CGIM 2002), Hawaii, USA, pp. 144-149, 2002.
- [17] P. C. A. Montenegro, M. Gattass, L. Velho, *Space Carving with a Hand-Held Camera*, SIBGRAP'2004 - International Symposium on Computer Graphics, Image Processing and Vision, 2004.
- [18] X. Meng, H. Li, Z. Hu, *A New Easy Camera Calibration Technique Based on Circular Points*, Proceedings of the British Machine Vision Conference, Bristol, UK, pp. 496-501, 2000.
- [19] Z. Zhang, *A New Multistage Approach to Motion and Structure Estimation: From Essential Parameters to Euclidean Motion Via Fundamental Matrix*, Technical Report, No. 2910, INRIA - Institut National de Recherche en Informatique et en Automatique, Sophia-Antipolis Research Unit, France, 1996.
- [20] Z. Zhang, *A Flexible New Technique for Camera Calibration*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 11, pp. 1330-1334, 2000.